Title: A Tool for Transformation of PDF to Text
Author: Jonáš Bujok
Department: Institute of Formal and Applied Linguistics (32-UFAL)
Supervisor: Mgr. Jan Raab, Institute of Formal and Applied Linguistics (32-UFAL)

Abstract: In this thesis we described an extraction procedure of text information from PDF (Portable Document Format) files. Thesis is focused mainly on middle-Europe languages. We designed, described and implemented program for this purpose. Besides the program and it's description the thesis contains information about PDF format object structure, it's syntax and logic necessary for proper understanding of text searching principles in PDF file. We also discussed filters, fonts and all other PDF Objects that the program need to process. This thesis also deals with methods and possibilities of improving program's functionality, speed, memory usage, reliability an universality of usage.